data - EDA

2026-02-03

PRIVACY WARNING: Sensitive columns detected

Overview

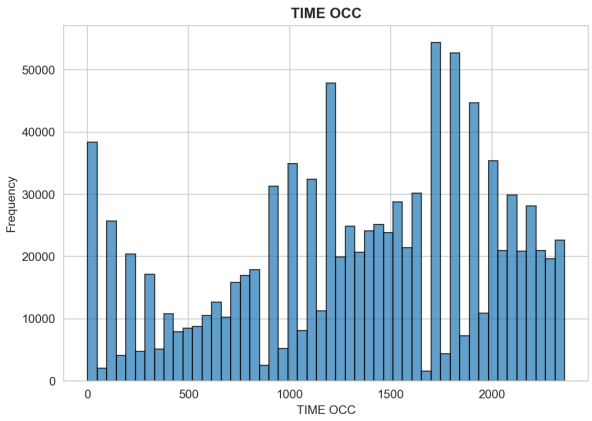
- Rows: 1,004,991
- Columns: 28
- Duplicates: 0
- High Missing: Weapon Used Cd, Weapon Desc, Crm Cd 2, Crm Cd 3, Crm Cd 4
- Dates: Date Rptd, DATE OCC

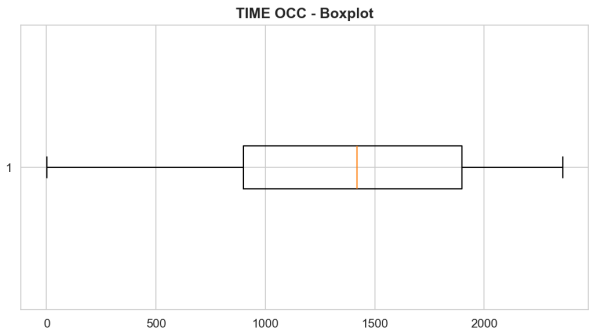
Numeric Analysis

TIME OCC

Min: 1.00 | Max: 2359.00 | Mean: 1339.90 | Median: 1420.00 | Std: 651.06

Outliers: 0 (0.0%)

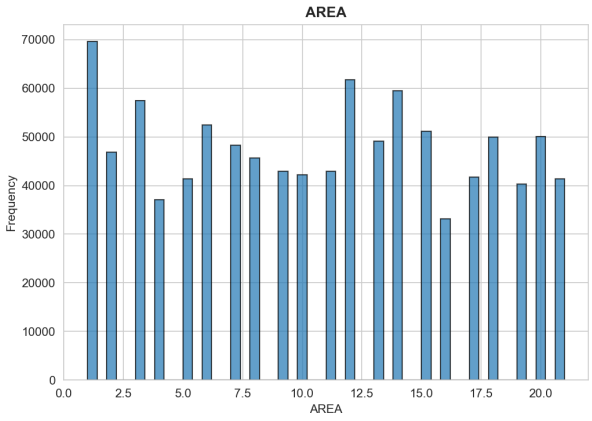


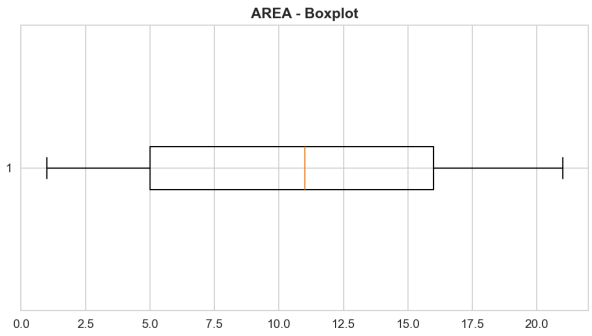


AREA

Min: 1.00 | Max: 21.00 | Mean: 10.69 | Median: 11.00 | Std: 6.11

Outliers: 0 (0.0%)

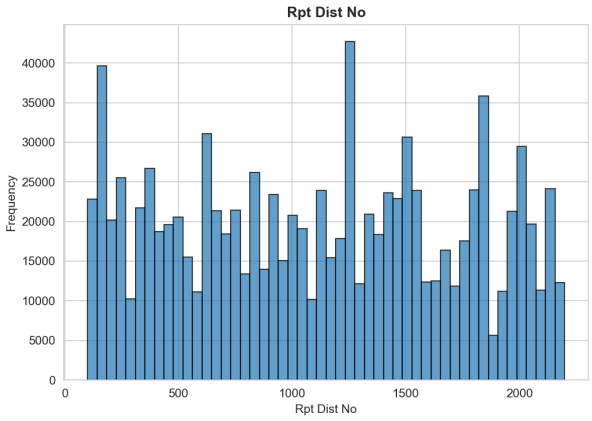


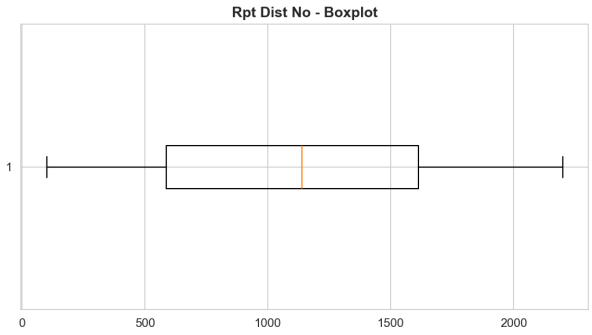


Rpt Dist No

Min: 101.00 | Max: 2199.00 | Mean: 1115.63 | Median: 1139.00 | Std: 611.16

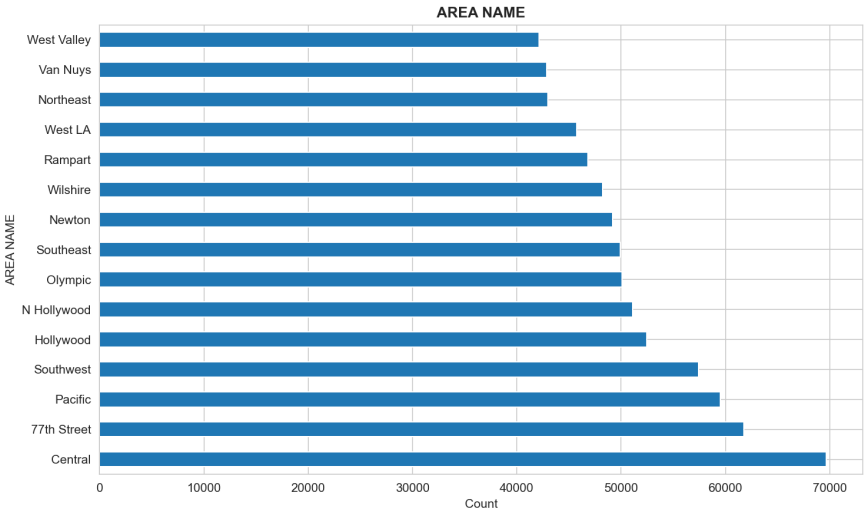
Outliers: 0 (0.0%)


Rpt Dist No
Frequency
40000
35000
30000
25000
20000
15000
10000
5000
0
0
500
1000
1500
2000
Rpt Dist No

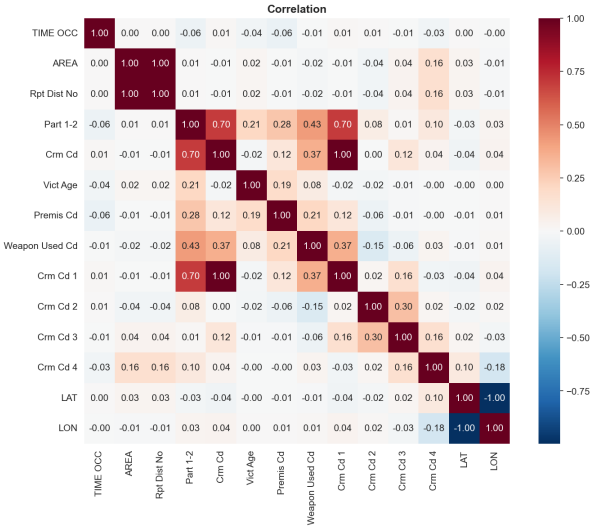

Rpt Dist No - Boxplot
1
0
500
1000
1500
2000

Categorical: AREA NAME

Value	Count	%
Central	69,670	6.9
77th Street	61,758	6.1
Pacific	59,514	5.9
Southwest	57,441	5.7
Hollywood	52,429	5.2
N Hollywood	51,107	5.1
Olympic	50,071	5.0
Southeast	49,936	5.0
Newton	49,177	4.9
Wilshire	48,239	4.8



Correlation



Key Insights

1. 1,004,991 rows x 28 cols, 80.1% complete
2. No duplicates - good quality
3. 6 columns >50% missing: Weapon Used Cd, Weapon Desc, Crm Cd 2
4. Crm Cd 2: 20.0% outliers
5. AREA NAME: 21 unique values, evenly distributed
6. Perfect correlation (1.00): Crm Cd and Crm Cd 1 are likely duplicates/derived
7. Date Rptd: 1981 day span
8. WARNING: 1 sensitive columns detected

Limitations

- Missing data in 6 columns may bias results
- Dataset represents specific time/location scope
- 1 columns may contain sensitive information

Auto-generated. All statistics verified.